STATISTICAL ANALYSIS PLAN

ALK9072-A306

Study Title:	A Phase 3b, Multicenter, Randomized, Double-blind Study to Evaluate the Efficacy and Safety of Aripiprazole Lauroxil or Paliperidone Palmitate for the Treatment of Schizophrenia in Subjects Hospitalized for Acute Exacerbation
Document Status:	Final Version
Document Date:	05 Feb 2019
Based on:	ALK9072-A306 Protocol v 2.0 (incorporates Amendment 1.0), Date: 04 Jan 2018, Original Protocol: 24 Aug 2017
Sponsor:	Alkermes, Inc. 852 Winter Street Waltham, MA 02451 USA

CONFIDENTIAL

Information and data in this document contain trade secrets and privileged or confidential information, which is the property of Alkermes, Inc. No person is authorized to make it public without the written permission of Alkermes, Inc. These restrictions or disclosures will apply equally to all future information supplied to you that is indicated as privileged or confidential. This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Abbreviation or Term	Explanation or Definition
AE	Adverse event
AIMS	Abnormal Involuntary Movement Scale
ATC	Anatomical Therapeutic Chemical [classification system]
AL	Aripiprazole lauroxil
AL-NCD	Aripiprazole lauroxil NanoCrystal Dispersion
ANCOVA	Analysis of covariance
BARS	Barnes Akathisia Rating Scale
BMI	Body mass index
CGI-I	Clinical Global Impression - Improvement
CGI-S	Clinical Global Impression - Severity
CI	Confidence interval
C-SSRS	Columbia Suicide Severity Rating Scale
C-VISA	Clinical-Validation Inventory for Study Admission
CYP	Cytochrome P450
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EC	Excited Component
ECG	Electrocardiogram
eCRF	Electronic case report form
EOT	End of treatment
EPS	Extra pyramidal symptoms
ESS	Epworth Sleepiness Scale
EQ-5D	EuroQOL-5D
ET	Early termination
FAS	Full analysis set
HbA1C	Hemoglobin A1c
IM	Intramuscular

Abbreviation or Term	Explanation or Definition
IRB	Institutional Review Board
ISR	Injection site reactions
LAI	Long-acting injectable
LOCF	Last observation carried forward
IWQOL	Impact of Weight on Quality of Life
MedDRA	Medical Dictionary for Regulatory Activities
MINI v 7.0.2	Mini-International Neuropsychiatric Interview-7.0.2 for Schizophrenia and Psychotic Disorder Studies
MSQ	Medication Satisfaction Questionnaire
PANSS	Positive and Negative Syndrome Scale
PBO	Placebo
PCS	Potentially clinically significant
PFS	Prefilled syringe
PK	Pharmacokinetic
PP	Paliperidone palmitate
PT	Preferred term
KM	Kaplan-Meier
QTcB	QT corrected with Bazett formula
QTcF	QT corrected with Fridericia formula
Q-LES-Q-SF	Quality of Life Enjoyment and Satisfaction Questionnaire - Short Form
RDQ	Readiness for Discharge Questionnaire
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Simpson-Angus Scale
SD	Standard deviation
SMQ	Standardized MedDRA queries
SOC	System organ class
TEAE	Treatment-emergent adverse event
UKU-SERS-PAT	Udvalg for Kliniske Undersøgelser Side Effect Self-Rating Scale – Patient

Abbreviation or Term	Explanation or Definition
ULN	Upper limit of normal
VAS	Visual analogue scale
WHO	World Health Organization
WoRQ	Work Readiness Questionnaire

1. INTRODUCTION

This statistical analysis plan (SAP) describes the statistical methods to be used for analyses and presentation of safety and efficacy data for the study ALK9072-A306. This document has been prepared based on Alkermes ALK9072-A306 Study Protocol Amendment 1.0 (dated 04 Jan 2018).

1.1. Study Objectives

1.1.1. Primary Objective

The primary objective of this study is to evaluate the efficacy of the aripiprazole lauroxil NanoCrystal® Dispersion (AL-NCD) initiation regimen (30 mg oral aripiprazole + 662 mg AL-NCD intramuscular [IM]) followed by aripiprazole lauroxil (AL) 1064 mg IM during the first 4 weeks of treatment of patients hospitalized for an acute exacerbation of schizophrenia.

1.1.2. Secondary Objectives

- To compare the efficacy of AL-NCD initiation regimen followed by AL 1064 mg IM with paliperidone palmitate initiation dosing (PP; Invega® Sustenna®; 234 mg + 156 mg IM) during the first 4 weeks of treatment of patients hospitalized for an acute exacerbation of schizophrenia
- To evaluate the safety, efficacy, and tolerability of AL 1064 mg IM after 6 months of treatment with AL for schizophrenia
- To compare the safety, efficacy, and tolerability of AL 1064 mg IM with monthly PP 156 mg IM after 6 months of treatment with AL or PP for schizophrenia

1.1.3. Exploratory Objectives

To characterize subject and caregiver centered outcomes, such as quality of life, work readiness, satisfaction with medication, resource utilization, and caregiver burden after 6 months of treatment with AL or PP for schizophrenia.

1.2. Summary of the Study Design

This is a multicenter, randomized, double-blind study evaluating the efficacy and safety of AL and PP in approximately 180 subjects experiencing an acute exacerbation of schizophrenia. In total, subjects will participate for approximately 26 weeks, including up to 1 week of inpatient screening and 25 weeks of treatment (which includes an initial 2-week inpatient stay). Potential subjects will be evaluated for eligibility according to the inclusion and exclusion criteria at a Screening visit (up to a 7 day period) and Baseline visit prior to randomization. Prior antipsychotic medications should be discontinued after screening upon inpatient admission. If a subject does not have historical exposure to aripiprazole and/or risperidone and paliperidone, test doses with aripiprazole and/or risperidone will be administered during the first 2 days of inpatient stay during the Screening period, prior to randomization, to assess tolerability.

On Day 1, qualified subjects will be randomized to either the ALtreatment group or the PP treatment group. Subjects will remain in the inpatient study unit during the Screening period and for at least 2 weeks after administration of the first dose of study drug on Day 1. Subjects will be discharged from the inpatient unit upon assessment as clinically stable and appropriate for discharge as determined by the Investigator. Subjects will receive their final injection of study drug on Week 21 (Day 148) and will be considered to be on treatment until Week 25 (Day 176).

Subjects meeting all eligibility criteria will be randomized to one of two treatment groups, as described in Table 1.

Table 1: Treatment Groups

Treatment Group	Aripiprazole lauroxil (AL)	Paliperidone palmitate (PP)
Day 1	662 mg AL-NCD (IM gluteal) + PBO (IM deltoid) + 30 mg oral aripiprazole	PP 234 mg (IM deltoid) + PBO (IM gluteal) + oral PBO
Day 8	AL 1064 mg (IM gluteal) + PBO (IM deltoid)	PP 156 mg (IM deltoid) + PBO (IM gluteal)
Days 36, 92, and 148	PBO (IM gluteal)	PP 156 mg (IM gluteal)
Days 64 and 120	AL 1064 mg (IM gluteal)	PP 156 mg (IM gluteal)

Abbreviations: AL=aripiprazole lauroxil; AL-NCD=aripiprazole lauroxil NanoCrystal Dispersion; IM=intramuscular; PBO=placebo; PP=paliperidone palmitate

To minimize potential response bias based on antipsychotic medication exposure history, randomization will be stratified by the status of prior exposure to aripiprazole or risperidone/ paliperidone. There are two stratification levels, as described in Table 2. After the stratification level (either Level 1 or Level 2) is determined for the subject, he or she will be randomized to one of two treatment groups in a 1:1 ratio. Stratification will ensure a balance of the two treatment groups within each stratification level.

Table 2: Stratification Factors

		Prior Risperidone/Paliperidone Exposure	
		Yes	**No**
Prior Aripiprazole Exposure	**Yes**	Level 1	Level 2
	No	Level 2	Level 1

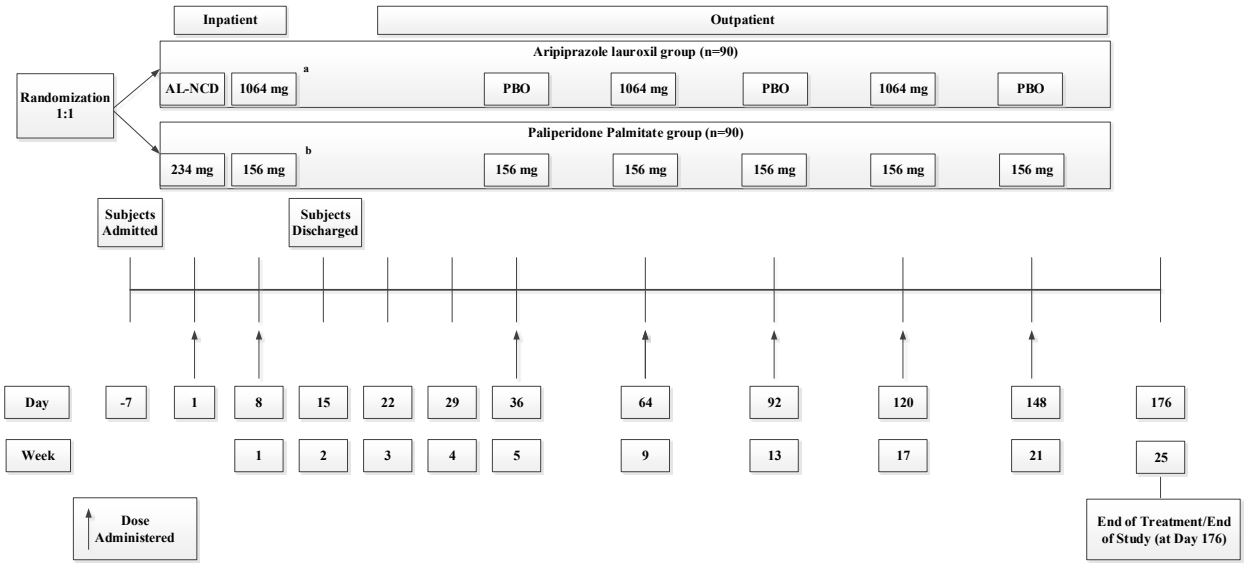
Safety and tolerability assessments will include adverse event (AE) monitoring, clinical laboratory testing, vital signs, body weight, 12-lead electrocardiogram (ECG) monitoring, abnormal movement scales (Abnormal Involuntary Movement Scale [AIMS], Barnes Akathisia Rating Scale [BARS], and Simpson-Angus Scale [SAS]), Columbia Suicide Severity Rating Scale (C-SSRS), injection site reactions (ISRs), the Udvalg for Kliniske Undersøgelser Side Effect Self-Rating Scale - Patient (UKU-SERS-PAT) − Sexual Side Effects Subscale, and Epworth Sleepiness Scale (ESS).

Efficacy assessments will include Positive and Negative Syndrome Scale (PANSS), Clinical Global Impression–Severity (CGI-S), and Readiness for Discharge Questionnaire (RDQ).

Subject or caregiver-centered assessments will include Quality of Life Enjoyment and Satisfaction Questionnaire - Short Form (Q-LES-Q-SF), modified Medication Satisfaction Questionnaire (MSQ), the Work Readiness Questionnaire (WoRQ), Handwriting Movement kinematics, Wrist actigraphy, and Resource Utilization Form. Caregivers who meet criteria will also complete the Burden Assessment Scale.

A schematic of the study design is provided in Figure 1.

Figure 1: Study Design Schematic



Abbreviations: AL-NCD=aripiprazole lauroxil NanoCrystal Dispersion; PBO=placebo
[a] Subjects randomized into the AL Treatment Group will receive IM PBO (deltoid) on Days 1 and 8 and 30 mg oral aripiprazole on Day 1
[b] Subjects randomized into the PP Treatment Group will receive IM PBO (gluteal) on Days 1 and 8 and oral PBO on Day 1

2. SAMPLE SIZE CONSIDERATION

No formal sample size calculations have been performed. The sample size of 180 subjects is based on practical, clinical considerations.

3. DATA ANALYSIS

3.1. General Statistical Methodology

Baseline for efficacy or safety analysis is defined as the last non-missing assessment before the first dose of study drug in ALK9072-A306 study on Day 1, and it will be used for all efficacy and safety analysis unless specified otherwise.

In general, descriptive statistics: n, mean (±standard deviation [SD]), median, minimum, and maximum, for continuous variables and number and percentage of subjects in each category for categorical variables will be provided.

All statistical tests and confidence intervals (CIs), unless stated otherwise, will be 2-sided and will be set at an alpha level of 0.05.

All source data will be presented as subject data listings.

3.2. Definitions of Analysis Populations

3.2.1. Safety Population

The Safety population will include all subjects who received at least one dose of study drug (AL injection, AL-NCD injection, 30 mg oral aripiprazole, oral placebo (PBO), PBO injection, or PP injection). Safety analyses will be based on the Safety population.

3.2.2. Efficacy Population

The Full Analysis set (FAS) will include all subjects in the Safety population who have at least one postbaseline assessment of PANSS. Efficacy analyses will be based on the FAS. Subjects identified as duplicate subjects are not included in the FAS Populations.

3.3. Disposition

The number and percentage of subjects completing or prematurely discontinuing the study including reasons for discontinuation will be summarized by treatment group and overall for the following:

- Subjects who enrolled in the study
- Subjects in the Safety Population
- Subject in the Efficacy Population
- Subjects who completed the 4-week Treatment period
- Subjects who completed the entire Treatment period
- Subjects who discontinued treatment along with reasons for discontinuation

3.4. Protocol Deviation

Subjects with major protocol deviations in the following categories will be summarized by treatment group and overall. A supportive listing will be provided as well.

- Did not meet the inclusion/exclusion criteria
- Received prohibited medications
- Lack of adherence with study medication, as defined by subjects taking 1 injection less than they were supposed to receive. A subject who missed the injection by > 2 weeks will also be considered as lack of adherence.
- Dosing error
- Others

3.5. Demographics and Baseline Characteristics

Demographics and baseline characteristics such as gender, age, race, ethnicity, weight, body mass index (BMI), prior antipsychotic medication, CYP2D6 phenotype, PANSS total and subscale scores, and CGI-S will be summarized by treatment group and overall for the Safety population.

3.6. Prior and Concomitant Medication

Prior medications are defined as medications taken prior to the first dose of study drug in ALK9072-A306 study. Concomitant medications are defined as medications taken on or after the first dose of study drug in ALK9072-A306 study. All medications will be coded using the World Health Organization Drug Dictionary (WHO-DD Enhanced + Herbal [version: March 2016 or higher]).

Concomitant medications taken during the Treatment period will be summarized by the preferred drug name for the Safety population. All reported medications (including prior medicatios and those initiated after the last dose of study medication) will be included in listing. For the summary table, if a subject has taken a concomitant medication more than once, the subject will be counted only once for that medication.

3.7. Treatment Adherence Rate and Duration of Study Drug Administration

3.7.1. Treatment Adherence Rate

All dosing and exposure information will be included in supportive listings. Cumulative number of injections by treatment group and overall for the safety population will be summarized.

3.7.2. Duration of Study Drug Administration

The duration of study administration will be defined as the last injection date minus the first dosing date + 30 days.

3.8. Efficacy Analyses

3.8.1. General Considerations

All statistical analysis will be performed at the 5% significance level. All CI will be 2-sided 95% confidence intervals. Efficacy analysis will be carried out using the FAS population. PANSS total and subscale scores, CGI-S, PANSS response analyses and body weight will be based on both observed data and the last observation carried forward (LOCF) imputation, that is, the last observed non-missing postbaseline value will be carried forward for missing postbaseline assessments.

In general, continuous and categorical endpoints will be summarized using descriptive statistics by treatment group and overall.

3.8.1.1. Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline in PANSS total score at Week 4 (Day 29).

3.8.1.2. Primary Analysis

The primary analysis is to compare the PANSS total score at Week 4 (Day 29) vs the total score at baseline within each treatment group separately. The two statistical hypotheses to be tested for the primary efficacy endpoints are:

1. For AL group: H_0: $d_{al\_week4} = 0$ versus H_a: $d_{al\_week4} \neq 0$

2. For PP group: H_0: $d_{pp\_week4} = 0$ versus H_a: $d_{pp\_week4} \neq 0$,

where d_{al_week4} is the change from baseline in PANSS total score at Week 4 (Day 29) for AL treatment group, and d_{pp_week4} is the change from baseline in PANSS total score at Week 4 for PP treatment group.

One-sample t-test will be used to assess the change from baseline in the PANSS total score at Week 4 (Day 29) against no improvement for the AL and PP treatment groups separately based on the observed data without imputation. Each analysis will be a 2-sided test with p–value = 0.05. No adjustments for multiplicity will be made for the primary analysis.

3.8.1.3. Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

- The change from baseline in PANSS total score at Week 4 (Day 29), Week 9 (Day 64) and Week 25 (Day 176).
- PANSS response by treatment group at each visit. PANSS response is defined as PANSS total score >=30% improvement from baseline.

3.8.1.4. Secondary Analysis

The first secondary analysis will be within treatment group comparison.The change from baseline in the PANSS total score at Week 9 (Day 64) and Week 25 (Day 176) will be tested against no improvement using the same approach outlined for the primary efficacy endpoint.

The second secondary analysis will be between treatment group comparison. Separate mixed models repeated measures (MMRM) will be used to analyze the results from baseline though either: Week 4 (Day 29), Week 9 (Day 64) or Week 25 (Day 176). The analyses will be based on the observed data. The model will include treatment group, visits, interaction term of treatment group and visit, baseline PANSS total score, stratification factor, and the pooled study sites as covariates. The model will assume an unstructured variance-covariance matrix for within-subject variability. If the model cannot converge due to the unstructured matrix, the heterogeneous autoregressive of order 1 will be used.

The change from baseline in PANSS total score will be plotted for each treatment group from the MMRM.

The third secondary analysis will be between treatment group comparison for PANSS response. PANSS response at each visit based on both the observed data and LOCF will be summarized by treatment group at each visit. A logistic regression model will be used to compare the two treatment groups at Week 4 (Day 29), Week 9 (Day 64) and Week 25 (Day 176); the model will include the treatment group, baseline PANSS total score, stratification factor, and the pooled study sites as covariates.

Additional imputation approaches will be explored.

3.8.1.5. Other Efficacy Endpoints

The observed values and change from baseline in PANSS subscale scores and CGI-S will be summarized. The change from baseline for the other efficacy endpoints will be analyzed using an analysis of covariance (ANCOVA) model with LOCF and the same MMRM as summarized above for the primary and secondary endpoints.

The change from baseline in PANSS subscale scores, and CGI-S will be plotted for each treatment group from the MMRM based on observed data, as well as from the ANCOVA model based on LOCF data.

Time from randomization to Readiness for Discharge will be analyzed using a Kaplan-Meier method. The log rank test will be used to compare two treatment groups.

Time to treatment discontinuation will be analyzed similarly using the Kaplan-Meier method by treatment group. Both, all cause discontinuation and medication related discontinuation (defined as due to AE or lack of efficacy) will be summarized. For prematurely discontinued subjects, time to discontinuation is defined as time from the date of first dose of study drug to the date of discontinuation. Subjects who compeleted the study will be considered censored at the end of treatment (EOT) visit. For medication related discontinuation analysis, early discontinued subjects due to other reasons will be considered censored at the discontinuation date.

3.8.2. Exploratory Analyses of Other Subject or Caregiver-Centered Assessments

3.8.2.5. Quality of Life Enjoyment and Satisfaction Questionnaire - Short Form

The Q-LES-Q-SF is a 16-item self-reported measure used to assess the degree of enjoyment and satisfaction experienced by subjects in various areas of daily functioning. Responses are evaluated on a 5-point scale (“very poor” to “very good”), with higher scores indicating more enjoyment and satisfaction. The scoring of the Q-LES-Q-SF involves summing only the first 14 items to yield a raw total score. The last two items are not included in the raw total score, but are standalone items. The raw total score is transformed into a percentage maximum possible (% Maximum) score using the following formula: (raw total score −14)/56. Subjects will complete the Q-LES-Q-SF at Week 5 (Day 36) , Week 13 (Day 92), and Week 25 (Day 176).The total raw score, % maximum score, and scores of item 15, 16 at each visit will be summarized by treatment group.

3.8.2.6. Burden Assessment Scale

The Burden Assessment Scale is a 19-item scale administered to caregivers who are family or friends (ie, nonprofessional caregivers) that focuses on specific subjective and objective consequences of families caring for individuals with severe mental disorders. Respondents are required to indicate whether they have experienced each of the types of burden ‘Not at all,’ ‘A little,’ ‘Some,’ or ‘A lot’ in the past month. These are scored 1, 2, 3, and 4, respectively. A higher score indicates more perceived burden. The Burden Assessment Scale will be

administered to the designated caregiver at Screening, Week 9 (Day 64), and Week 25 (Day 176). The change from baseline in Burden score will be summarized by treatment group.

3.8.2.7. Modified Medication Satisfaction Questionnaire

The modified MSQ is a 3-item self-reported patient satisfaction questionnaire which assesses the level of patient satisfaction with medication. Subjects rate their satisfaction with their current medication, their preference for their current medication vs previous medications, and their opinion on the side effects of their current medication vs previous medications. Ratings are on a 5-point Likert scale. MSQ was assessed on Day 29 (Week 4), 64 (Week 9), 120 (Week 17), and 176 (Week 25). The number and percentage of response for the modified MSQ will be summarized by treatment group.

3.8.2.8. Work Readiness Questionnaire

The WoRQ is a questionnaire (7-item and a global assessment) used to assess a patient's ability to engage in socially useful activity, independent of work availability. The Investigator or designee will complete the WoRQ at Day 15 (Week 2), 29, (Week 4), 64 (Week 9), 120 (Week 17), 176 (Week 25). The global assessment (yes or no on ready to work) will be summarized by treatment group.

3.8.2.9. Resource Utilization

The number and percentage of subjects who had an emergency room visit, arrest, hospitalization (based on the serious adverse events [SAEs] that led to hospitalization), or have been incarcerated will be summarized by treatment group.

3.8.2.10. Handwriting Kinematics

A quantitative analysis of handwriting kinematics is used to evaluate effects of antipsychotic medication in patients with schizophrenia. The Investigator or designee will administer the handwriting tests to the subjects at screening, Week 2 (Day 15), Week 4 (Day 29), Week 9 (Day 64), and Week 17 (Day 120). The following endpoints related to extrapyramidal syndrome (EPS) will be summarized by treatment group.

- Stroke duration
- Stroke amplitude or size
- Peak and average velocity
- Frequency of secondary movements
- Average normalized jerk
- Number of acceleration peaks
- Pen pressure variability

Peak and average velocity and frequency of secondary movements (but not limited to) will be analyzed separately for a subset of patients with signs of EPS by treatment group. The subset of patients for this analysis will be selected based on the baseline SAS score of ≥5.

3.8.2.11. Wrist Actigraphy

Wrist actigraphy will be analyzed using the derived activity and sleep parameters collected during the two 14-days monitoring Cycles following the distribution of actigraph. The actigraph device will be distributed on Week 3 (Day 22) and Week 9 (Day 64). For each collection period, actigraphy parameters will include: total time sleep (min), sleep efficiency (%), sleep onsets (min), number of wakes after sleep onset, wakes after sleep onset (min), mean wake duration after sleep onset (min), deep rest time (min), deep rest sleep time (min), nap rest time (min), nap rest sleep time (min), day hour of peak activity, day mean activity in peak hour, day standard deviation of activity within peak hour, circadian rhythm period (min), circadian rhythm mesor, circadian rhythm amplitude.

Length of daytime activity will be calculated based on the the following rules:

1. Sleep onset is defined as the first recorded instance of sleep occurring at or after 9 pm and if the subject was asleep at 9 pm, using the latest transition from wake to sleep prior to 9 pm.
2. Sleep offset is defined as the last instance of transitioning from sleep to wake before 9am on the following day and if the person was still labelled as asleep at 9am, considering the next transition from sleep to wake as the sleep offset.
3. All activity that falls outside of this ‘sleep window’ is defined as daytime activity.

In each cycle, for each subject, the average of 14 days data will be calculated based on the available individual data. Average for the two monitoring cycles and for the combined total will be used in the summary statistics.

Wrist actigraphy parameters will be summarized by the treatment group, and by subgoups. Definitions of the subgroups are provided in Table 3.

Table 3: Subgroup Definitions

#	Target subjects	Criteria
A	Subjects with prominent negative symptoms	*Baseline* score of ≥ 4 on at least 3, or ≥ 5 on at least 2, of the 7 Negative PANSS Subscale Scores
B	Subjects with Lack of efficacy or psychiatric AEs *(treatment failure)*	Subjects with Lack of efficacy, OR Subjects with Psychiatric SAEs or AE leading to discontinuation (AEDC) of (PT): Schizophrenia Acute psychosis Suicidal behaviour Suicidal ideation Suicide attempt, OR Subjects who experienced worsening of psychosis defined as increase of PANSS (>30%) and CGI-S (2-point worsening) scores from baseline at any point
C	Subjects with Akathisia	BARS score of ≥ 3 at Week 4 (Day 29), OR Subjects with AEs of (PT) ongoing during the recording cycle: Akathisia Restlesness
D	Subjects with marked daytime sleepiness and sedation	ESS scale >16 at Week 4 (Day 29) and at Week 9 (Day 64) for Cycle 1 and Cycle 2, respectively, OR Subjects with AEs of (PT) ongoing during the recording cycle: Somnolence Sedation Hypersomnia Lethargy
E	Subjects on Benzodiazepines	Received the drugs with the following WhoDD codes during the recording cycle N05 PSYCHOLEPTICS N05CD Benzodiazepine derivatives N05CF Benzodiazepine related drugs
F	Subjects with Insomnia	Patients with ongoing Medical History of Insomnia (PT), OR Patients with AEs of Insomnia (PT) ongoing during the recording cycle
D	Severity of Psychiatric Illness	Subjects will be separated onto two subsets based on the CGI-S scale score on Day 22: mildly ill subgroup with the score of 3 and below, and

#	Target subjects	Criteria
		moderately ill subgroup – with the score of 4 and greater. Actigraphy parameters of the recording Cycle 1 (Day 22 to Day 36) are used in this table.

3.8.3. Multiple Comparison/Multiplicity

Primary endpoints for both treatment groups will each be tested at a full α level of 0.05. The study will be claimed positive when both of the the primary analyses are statistically significant. Only if both primary endpoints are statistically significant will the secondary endpoints be further tested at an alpha level of 0.05.

3.8.4. Pooling Sites

Approximately 16 US sites are involved in this study. To insure the precision of the estimation, a minimum of 10 subjects per treatment group within each site may be pursued by pooling the actual sites as following: rank the sites by the number of subjects at end of study from largest to smallest. If the smallest site has less than 10 subjects in one treatment group, pool this site with the second smallest site. If the pooled site has at least 10 subjects per treatment group, stop pooling. Otherwise, continue pooling with the third smallest site until each treatment group has at least 10 subjects within pooled site. However, in order to avoid a dominant center due to pooling, the pooling shall be stopped if the number of subjects within pooled site exceeds 20% of the total number of subjects.

3.9. Safety Analysis

3.9.1. General Considerations

All safety endpoints will be summarized by treatment group and overall for the Safety Population.

3.9.2. Adverse Events

Incidence of treatment emergent AEs will be analyzed as a safety endpoint. Adverse events will be coded by System Organ Class (SOC) and Preferred Term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA) Version 21.1 or higher. The verbatim term will be included in the AE listings.

An AE (coded by PT) will be considered as treatment emergent AE (TEAE) if the event is newly occurring or worsening on or after the date of first dose of study drug in this study.

An overview table, including number of subjects with TEAEs, AEs leading to study discontinuation (AEDCs), SAEs, study drug related TEAEs, will be provided.

The number and percentage of subjects reporting TEAEs during the Treatment period will be presented by treatment group and overall for the following categories:

- SOC and PT
- PT, and including the following subset:
  - TEAEs experienced by ≥5% of subjects (in any group)
- SOC, PT, and severity
- SOC, PT for severe TEAEs
- SOC, PT, and relationship
- SOC, PT for study drug related TEAEs
- SOC, PT for SAE

If the same PT occurred more than once for the same subject, the subject will be counted only once for that PT using the most severe and most related occurrence for the summarization by severity and by relationship to the study drug.

In additional, the following AE tables will be summarized by the cumulative 4-week and 9-week periods separately:

- Overall summary of AE
- All AE by SOC and PT
- SAE by SOC and PT
- AEDC by SOC and PT

The onset and duration of following individual AEs will be displayed graphically:

- Akathisia (including PTs of Akathisia, Extrapyramidal disorders, Hyperkinesia, Hyperkinesia neonatal, Motor dysfunction, Movement disorder, Psychomotor hyperactivity, Restlessness) for each treatment group
- Sedation (including PTs of Sedation, Somnolence, Hypersomnia, Lethargy) on benzodiazepines vs. non-benzodiazepines for each treatment group
- Hypotension/orthostasis (including PTs of Hypotension, Blood pressure orthostatic, Orthostatic hypotension, Presyncope, Syncope, Dizziness, Tachycardia) in antihypertensive vs. not for each treatment group
- ISRs for each treatment group

3.9.3. Death, Serious and Other Significant Adverse Events

The number and percentage of subjects who have SAE and AEDCs will be summarized by system organ class, preferred term, and treatment group and overall. Any deaths, subjects with serious AEs, and subjects who discontinue due to AEs will be listed.

3.9.3.1 Adverse Events of Special Interest

In addition, incidence of a selected subset of relevant AEs in this class of drugs will be summarized by preferred term and treatment group and overall for cumulative 4-week, 9-week and entire Treatment period separately. AEs will include the following:

- Akathisia
- Dyskinesia
- Dystonia
- Parkinson-like events
- Suicidal Ideation and Behavior

The selection of AEs per subset will be based on the PT and customized or Standardized MedDRA queries (SMQs). The list of PTs will be provided as an Appendix.

3.9.4. Clinical Laboratory Parameters

Laboratory parameters will be presented in conventional (ie, US) units. Only scheduled laboratory parameters will be included in the summaries, unless specified otherwise. All laboratory data, including those collected at unscheduled visits, will be included in the listings.

Laboratory results including baseline and change from baseline for the Safety Population during the Treatment period for chemistry and hematology parameters will be summarized by treatment group, overall, and by visit.

Clinical laboratory test values, scheduled or unscheduled, will be considered potentially clinically significant (PCS) if they meet PCS criteria listed in Table 4. The number and percentage of subjects who have postbaseline PCS clinical laboratory values will be summarized by treatment group and overall. The percentages will be calculated based on the number of subjects with non-PCS baseline value and have at least one postbaseline assessment. All PCS values including baseline PCS values will be included in supportive listings.

Shift tables for selected metabolic parameters (glucose, total cholesterol, LDL, HDL, tryglyceride, and HbA1c) and liver function tests will be presented. The criteria are summarized in Table 5, Table 6, and Table 7 .

In addition, the number (percentage) of subjects who had a prolactin value >1 × ULN, >2 × ULN, >3 × ULN at any post-baseline will be also summarized by gender, visit, and treatment group. Baseline prolactin will be summarized by prior treatment for each gender.

The selected lab parameters will be presented by figures. The prolactin and change from baseline over time will be plotted by gender and treatment group. The weight and vital signs and changes from baseline will be plotted over time.

Number of subjects who meet Hy's Law criteria (ALT or AST >3xULN, along with total Bilirubin >2xULN and a non-elevated ALP<2xULN) will be summarized by treatment group.

Table 4: Criteria of Potentially Clinically Significant (PCS) Abnormality for Selected Analytes

Parameters	Criteria
Chemistry	
Albumin	<2.5 g/dL
Alkaline Phosphatase (U/L)	≥3 × ULN
Alanine Aminotransferase (U/L)	≥3 × ULN
Aspartate Aminotransferase (U/L)	≥3 × ULN
Bilirubin, Total	≥2.0 mg/dL
Blood Urea Nitrogen (BUN)	>30 mg/dL
Creatinine Kinase (CK)	> 3 x ULN
Creatinine	≥ 2.0 mg/dL
Creatine phosphokinase (CPK), Total	≥ 3 x ULN
Cholesterol, Random	>300 mg/dL
Cholesterol, Fasting	≥240 mg/dL
Cholesterol, HDL Fasting	≤30 mg/dL
Cholesterol, LDL Fasting	≥160 mg/dL
Creatine Kinase (U/L)	≥3 × ULN
Creatinine	≥2.0 mg/dL
Glucose, Random	<50 mg/dL or ≥200 mg/dL
Glucose, Fasting	<50 mg/dL or ≥126 mg/dL
Potassium	<3 mmol/L or >5.5 mmol/L
HbA1c%	>6.5%
Lactate Dehydrogenase (U/L)	>3 × ULN
Prolactin by gender	>1 × ULN >2 × ULN >3 × ULN

Parameters	**Criteria**
Sodium	<130 mmol/L or >150 mmol/L

Table 4: Criteria of Potentially Clinically Significant (PCS) Abnormality for Selected Analytes (Continued)

Parameters	Criteria
Hematology	
Eosinophils	$>1.0 \times 10^3/\mu L$
Hematocrit (Female)	≤32%
Hematocrit (Male)	≤37%
Neutrophils, Absolute	$<1.5 \times 10^3/\mu L$
Platelets	$<75.0 \times 10^3$ cells/μL or $\geq 700.0 \times 10^3$ cells/μL
Leukocytes	$\leq 2.8 \times 10^3/\mu L$ or $\geq 16.0 \times 10^3/\mu L$

Table 5: Shifts Category from Baseline to Any Postbaseline for Selected Lipid Parameters

Total Cholesterol (fasting) mg/dL
Normal (<200) to High (≥240)
Borderline (≥200 and < 240) to High (≥240)
Normal/Borderline (<240) to High (≥240)
Normal (<200) to borderline/High (≥200)
Increase ≥40 mg/dL
LDL Cholesterol (fasting) mg/dL
Normal (<100) to high (≥160)
Borderline (≥100 and <160) to high (≥160)
Normal/borderline (<160) to high (≥160)
Normal (<100) to borderline/high (≥100)
Increase ≥30 mg/dL
HDL Cholesterol (fasting) mg/dL
Normal (≥40) to low (<40)
Decrease ≥20 mg/dL

Table 5: Shifts Category from Baseline to Any Postbaseline for Selected Lipid Parameters (Continued)

Total Cholesterol (fasting) mg/dL
Triglycerides (fasting) mg/dL
Normal (<150) to high (≥200)
Normal (<150) to very high (≥500)
Borderline (≥150 and <200) to high (≥200)
Borderline (≥150 and <200) to very high (≥500)
Normal/borderline (<200) to high (≥200)
Normal/borderline (<200) to very high (≥500)
Normal (<150) to borderline/high/very high (≥150)
Increase ≥50 mg/dL

Table 6: Shift Category from Baseline to Any Postbaseline in Glucose and HbA1c

Serum glucose (fasting) mg/dL
Normal (<100) to High (≥126)
Impaired (≥100 and <126) to High (≥126)
Normal/Impaired (<126) to High (≥126)
Increase ≥10 mg/dL
HbA1c %
Shift from baseline (<5.7%) to postbaseline ≥5.7%
Shift from baseline (<5.7%) to postbaseline ≥5.7% and <6.5%
Shift from baseline (<5.7%) to postbaseline ≥6.5%

Table 7: Shift Category from Baseline to Any Postbaseline in Liver Function Test

Alanine Aminotransferase (ALT) (U/L)
Shift from Normal to ≥3 x ULN
Shift from Normal to ≥5 x ULN
Shift from Normal to ≥10 x ULN
Aspartate Aminotransferase (AST) (U/L)
Shift from Normal to ≥3 x ULN
Shift from Normal to ≥5 x ULN
Shift from Normal to ≥10 x ULN
Bilirubin, Total (mg/dL)
Shift from Normal to >1 x ULN
Shift from Normal to ≥2 x ULN

3.9.5. Vital Signs, Body Weight, and Electrocardiograms

3.9.5.1. Vital Signs

Descriptive statistics for vital signs and changes from baseline values at each scheduled time point will be presented by treatment group for the Treatment period.

Vital sign values will be considered PCS if they meet both the observed-value criteria and the change-from-baseline criteria listed in Table 8. The number and percentage of subjects with PCS postbaseline values will be tabulated by treatment group. The percentages will be calculated based on the number of subjects with non-PCS baseline values and at least one postbaseline assessment. The numerator will be the number of subjects with non-PCS baseline values and at least one postbaseline PCS value. A supportive listing of subjects with PCS postbaseline values will be provided.

All vital signs will be presented in the subject data listing.

Orthostatic hypotension can be defined in two ways. The first is defined as a fall in systolic blood pressure of at least 20 mmHg and a fall in the diastolic blood pressure of at least 10 mmHg upon standing from supine. The other Orthostatic hypotension is defined as a fall in systolic blood pressure of at least 30 mmHg upon standing from supine.

Orthostatic tachycardia is defined as a heart rate increase of 30 beats per minute (bpm) or more upon standing from supine, or over 120 bpm upon standing.

The number and percentage of subjects with orthostatic hypotension or orthostatic tachycardia occurring at any postbaseline visit will be summarized for by treatment group.

Table 8: Criteria for Potentially Clinically Significant Blood Pressure or Pulse Rate

Parameter	Criteria
Supine Systolic Blood Pressure	≤90 and decrease ≥20 mm Hg ≥180 and increase ≥20 mm Hg
Supine Diastolic Blood Pressure	≤50 and decrease ≥15 mm Hg ≥105 and increase ≥15 mm Hg
Supine Heart Rate	≤50 and decrease ≥15 bpm ≥120 and increase ≥15 bpm

3.9.5.2. Weight and Body Mass Index

Weight (kg), BMI (kg/m^2) (baseline and change from baseline) will be summarized by treatment group and overall.

In addition, number and percentage of subjects with weight change values considered as PCS occurring at any post-baseline visit will be summarized by treatment group and overall. Criteria for PCS are presented in Table 9. The percentages will be calculated relative to the number of subjects in the Safety Population with at least one post-baseline value. A supportive listing will be provided for subjects with PCS values. The PCS increase in weight will also be summarized by baseline BMI category:

- Underweight [<18.5 kg/m^2];
- Normal [≥18.5 to <25 kg/m^2];
- Overweight [≥25 to <30 kg/m^2]; and
- Obese [≥30 kg/m^2].

Table 9: Criteria for Potentially Clinically Significant Changes from Baseline in Body Weight

Parameter	Criteria
Body Weight	Decrease from Baseline ≥7% Increase from Baseline ≥7%

3.9.5.3. Electrocardiograms

Descriptive statistics for ECG parameters (heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc interval) at baseline and change from baseline values at each scheduled assessment timepoint and at the end of the Treatment period will be presented by treatment group for the Treatment period. QTc interval will be calculated using both Bazett (QTcB = QT/

$(RR)^{1/2}$) and Fridericia (QTcF = QT/ $(RR)^{1/3}$) corrections; if RR is not available, it will be replaced with 60/HR in the correction formula.

Electrocardiogram parameter values are considered PCS if they meet or exceed the higher-limit PCS criteria listed in Table 10. The number and percentage of subjects with PCS postbaseline ECG values will be tabulated by treatment group. The percentages will be calculated based on the number of subjects with non-PCS baseline values and at least one postbaseline assessment. The numerator is the number of subjects with non-PCS baseline values and at least one postbaseline PCS value. A supportive listing of subjects with PCS values will be provided.

Table 10: Criteria for Potentially Clinically Significant (PCS) QTcF

Parameter	Criteria
QTcF	>450 to ≤480 msec
	>480 to ≤500 msec
	>500 msec
	Change from baseline >30 to ≤60 msec
	Change from baseline >60 msec

3.9.6. Abnormal Movement Scales

Extra pyramidal symptoms will be evaluated as AEs and also as assessed by abnormal movement scales. Abnormal movement scales will include the following: AIMS, BARS, and SAS.

For all abnormal movement scales, total scores and subscale scores will be summarized by treatment group at each visit for the absolute value and for changes from baseline.

Number and percentage of subjects meeting the criteria for treatment emergent Parkinsonism (SAS total score>3), for treatment emergent akathisia (BARS global clinical assessment of akathisia score ≥2), for treatment emergent dyskinesia (AIMS score ≥3 on any of the first 7 items, or a score ≥2 on two or more of any of the first 7 items) at any postbaseline visit will be summarized by treatment group.

A listing will be provided for every abnormal movement scale. Listing for treatment emergent EPS will be provided.

3.9.7. Injection Site Reactions

All ISR will be reported as AEs. Injection site reaction AE preferred terms (injection site pain, injection site erythema, injection site swelling, etc.) will be summarized by treatment group and overall.

3.9.8. Columbia-Suicide Severity Rating Scale

The C-SSRS is a questionnaire used to measure the presence and intensity of suicidal ideation and behavior.

Suicidal behavior and suicidal ideation will be summarized for the Safety population. The number of subjects with suicidal ideation and suicidal behavior will be summarized by treatment group when applicable.

Supportive tabular display of subjects with all values will be provided.

Table 11: C-SSRS Categories for Analysis

Category	C-SSRS Item response is "YES"
Suicidal behavior	Preparatory acts or behavior Aborted attempt Interrupted attempt Actual attempt Complete suicide
Suicidal ideation	Wish to be dead Non-specific active suicidal thoughts Active suicidal ideation with any methods (not plan) without intent to act Active suicidal ideation with some intent to act, without specific plan Active suicidal ideation with specific plan and intent

3.9.9. Epworth Sleepiness Scale

The ESS is a self-reported questionnaire with 8 questions that provides a measure of a person's level of daytime sleepiness. The ESS assesses, on a 4-point scale (0–3), a person's usual chances of dozing off or falling asleep in 8 different situations or activities. This scale will be completed by the subject at Week 2 (Day 15), Week 3 (Day 22), Week 4 (Day 29), Week 9 (Day 64), Week 13 (Day 92), and Week 25 (Day 176). The sum of 8-item score will be summarized by treatment group.

3.9.10. UKU-SERS-PAT-Sexual Side Effects

The sexual functioning subscale from the UKU-SERS-PAT contains 7 items for males and 9 items for females. Each item is rated on a 4-point scale (0=none or doubtful, 1=present to a mild degree, 2=present to a moderate degree, and 3=present to a severe degree). This scale will be completed by the subject at the time points specified at Day 1, Week 4 (Day 29), Week 9

(Day 64), Week 17 (Day 120), and Week 25 (Day 176). The UKU score will be summarized by the following way:

- Total score and change from baseline will be summarized by treatment group and gender
- The number and percentage of subjects with ≥1 score on any item at each visit will be summarized by treatment group and gender (all items)
- The number and percentage of subjects with ≥1 score on each item at any visit will be summarized by treatment group and gender (each item reported separately)

3.9.11. Genotyping data

CYP2D6 phenotype will be summarized by treatment group and overall. The change from baseline in PANSS and CGI-S at Week 4 (Day 29), Week 9 (Day 64), and Week 25 (Day 176) will be analyzed by each phenotype status (extensive metabolizer, intermediate metabolizer, or poor metabolizer). In addition, the AE incidence rate will be also summarized by phenotype status and treatment group.

3.10. Pharmacokinetic/ Pharmacodynamic Data Analysis

Not applicable.

4. INTERIM ANALYSES

There was no interim analysis for this study.

5. CHANGES IN CONDUCT OR PLANNED ANALYSES FROM THE PROTOCOL

The ANCOVA with LOCF for the secondary efficacy endpoint (the change of PANSS total score from baseline at Week 4 [Day 29], Week 9 [Day 64], and Week 25 [Day 176]) was replaced by MMRM approach to better address the missingness assumption.

6. DEFINITIONS AND CONVENTIONS FOR HANDLING OF THE DATA

Dataset specifications will be provided in a separate document.

6.1. Analysis Visit Windows

Scheduled analysis visits are visits at scheduled timepoints as specified in the protocol (Table 1 Schedule of Visits and Assessments).

Scheduled analysis visits during the study period will be the same as the nominal visits collected in eCRF. There will be one valid value of assessment kept for each scheduled analysis visit in summary/analysis statistics.

Unscheduled visits are visits with data not collected on the scheduled time point. Unscheduled visits will not be used for by-visit summary/analysis statistics, unless specified otherwise.

All unscheduled visits as collected in eCRF will be included in listings.

Visit Day is calculated as date of visit – date of the first dose of study drug + 1 day.

Last postbaseline values are defined as the last valid postbaseline values collected for each subject during the Treatment period.

6.2. Handling of Partial Dates of Concomitant Medication

Partial start dates of prior and concomitant medications will be assumed to be the earliest possible date consistent with the partial date. Partial stop dates of prior and concomitant medications will be assumed to be the latest possible date consistent with the partial date. In the case of completely missing stop date, medication will be assumed to be ongoing.

6.3. Handling of Safety Data

All efforts should be made to obtain the missing information from the investigator. For C-SSRS, vital signs, laboratory testing (chemistry, hematology, urinalysis), and 12-lead ECGs, only observed data will be used for analyses, and missing data will not be imputed.

7. GENERAL STATISTICAL METHODOLOGY

In general, summary statistics (n, mean, SD, median, minimum, and maximum for continuous variables, and number and percentage of subjects in each category for categorical variables) will be provided by treatment group. All summary tables will be based on observed data, and missing values will not be imputed, unless otherwise indicated. Measurements collected from unscheduled visits or repeated assessments will not be included in the by-visit summary tables or figures, but will be included in the analyses for the PCS postbaseline values, and subject listings. Source data for the summary tables will be presented as subject data listings.

7.1. Reporting Precision

Summary statistics will be presented to the following degree of precision, unless otherwise specified:

Table 12: Degree of Precision

Statistics	Degree of Precision
Mean, Median, Confidence limit boundaries	One more than the raw data, up to 3 decimal places.
Standard deviation, Standard error	One more than the mean, up to 3 decimal places
Minimum, Maximum	The same as the raw data, up to 2 decimal places
P-value	Rounded to 3 decimal places and therefore presented as 0.xxx; *P*-values smaller than 0.001 as '<0.001'; P-values greater than 0.999 as '>0.999'.
Percentage	One decimal place. A percentage of 100% will be reported as 100%. Percentages of zero will be reported as 0.

Fractional numeric values will be presented with a zero to the left of the decimal point (for example, 0.12 – 0.30).

For weight, height, and BMI, one decimal place will be used for summary statistics, unless otherwise specified.

8. PROGRAMMING SPECIFICATIONS

Programming specifications will be provided in a separate document.

9. MOCK TABLES, LISTINGS AND FIGURES

Mock-up tables and listings will be provided in a separate document.

10. REFERENCES

1. Alkermes ALK9072-A306 Study Protocol Amendment Date: 04 Jan 2018